UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

CRT-575

MODELLING OF A DEREGULATED TRANSIT SYSTEM

by

Enrique Fernandez*
and
Patrice Marcotte**

* Pontificia Universidad Catolica de Chile, Santiago, Chile.

** Centre de recherche sur les transports, Université de Montréal, and Collège Militaire Royal de Saint-Jean, Saint-Jean-sur-Richelieu, Canada.

This research has been financed by grants from DIUC, FONDECYT (Chile) and IDRC, NSERC (Canada).

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Université de Montréal
Centre de recherche sur les transports - Publication #575
June 1988

Dépôt légal - 2e trimestre 1988

Abstract. We present two formulations of a model for finding equilibrium passengers and operators flows in a deregulated transit system where bus operators are free to choose the routes where they offer public transport services. Algorithms corresponding to each formulation are proposed.
Key words: equilibrium, transit networks, bus operators.

Résumé. Nous proposons deux formulations équivalentes pour un problème d'équilibre usagers-opérateurs dansun système de transport en commun déréglementé, dans lequel chaque opérateur est libre de choisir son trajet (ligne). Nous présentons des algorithmes correspondant à chaque formulation.
Mots–clés: equilibre, réseaux de transport en commun, opérateurs d'autobus.

1. Introduction.

In this paper we consider an urban public transport system served by buses that are individually owned and operated by small entrepreneurs under deregulated conditions. Given a set of lines where public transport can be offered, private operators freely decide on which of these lines to run their buses such that their net benefit be maximized.

Depending on the number of buses operating on each line, the length of the route operated and the congestion experienced by the vehicles, a service frequency will result which in turn will influence the demand on that line and therefore the operators benefits.

If the number of buses in the system is large, as is the case in many developing countries, an important part of the congestion effect will be induced by buses (in contrast with cars) thus affecting their operating costs and that of the private cars using the same network infrastructure.

The problem we consider is that of finding the equilibrium state of the passengers-operators system. In the operators subsystem, the level of service is determined through the operators decisions (route choices). The second subproblem consists in finding a bimodal (private car – transit users) equilibrium.

This model is a good representation of the present situation in cities in Chile in particular and in developing countries in general. In Santiago, the capital city of Chile, the daily demand for four million public transport trips is mainly satisfied by an eight thousand buses fleet operated by small entrepreneurs, with an average of 1.5 vehicle per entrepreneur.

2. Notation and basic modelling assumptions.

Let $G = (\mathcal{N}, \mathcal{A})$ denote a transportation network with node set $\mathcal{N}$ and link set $\mathcal{A}$. The set of links represents city streets and the set of nodes represents intersections and centroids where trips originate or terminate. We assume that

G serves as an infrastructure for both public (buses) and private (cars) transport modes. Most links in $\mathcal{A}$ will be simultaneously used by private cars and buses, and therefore congestion interactions will exist: Changes in car flows will influence bus operating costs and service levels, and reciprocally. We will denote:

W : set of network origin-destination (O-D) pairs.

w : an element of set W.

P : set of paths in G available to private cars.

P_w : set of feasible paths associated with O-D pair w.

p : an element of set P.

T : set of O-D demands for private transport.

T_w : private transport demand for O-D pair w.

f_a : car flow on link a.

h_p : car flow on path p.

c_a : generalized travel cost for car users, over link a.

C_p : generalized travel cost for car users, over path p.

$\{\delta_{ap}\}$: arc-path incidence matrix for private transport:
-takes a value of 1 if arc a belongs to path p,
-and 0 otherwise.

Using the preceding notation, we define the vectors:

$$f = \{f_a | a \in \mathcal{A}\}, \quad c = \{c_a | a \in \mathcal{A}\}, \quad C = \{C_p | p \in P\}.$$

Note that the following relations hold between demand and flow variables:

$$\text{Demand constraints:} \qquad T_w = \sum_{p \in P_w} h_p, \qquad \forall w \in W \tag{1}$$

$$\text{Flow conservation:} \qquad f_a = \sum_{p \in P} \delta_{ap} h_p, \qquad \forall a \in \mathcal{A} \tag{2}$$

In addition we consider a transit network $\overline{G}$ consisting of a subset of nodes $\overline{\mathcal{N}}$ of $\mathcal{N}$, and a set of transit lines $\mathcal{L}$. Each transit line is defined as a sequence of nodes where passengers can board and alight. A transit line can also be described by a sequence of links of $\mathcal{A}$. The transit network is a representation of transit services available to customers. We will make use of the following notation:

R : set of feasible transit routes in $\overline{G}(\overline{\mathcal{N}}, \mathcal{L})$.

R_w : set of feasible transit routes for O-D pair w, per unit of time.

r : element of the set R.

$\overline{T}_w$: transit demand per unit of time, associated with O-D pair w.

$\overline{\mathcal{A}}$: set of links of network G ($a \in \mathcal{A}$) used by transit lines $l \in \mathcal{L}$.

$\overline{f}_a$: : total transit passengers flow on link a, per unit of time.

$\overline{f}^l$: total number of passengers taking line l per unit of time.

$\overline{f}_a^l$: number of passengers going on line l at link a, per unit of time.

$\widetilde{f}_a^l$: number of passengers boarding line l at link a, per unit of time.

$\widetilde{f}_{aj}^l$: number of passengers boarding line l at link a, and alighting at destination or transfer node j.

φ^l : service frequency on line l.

φ_a : total number of buses passing through link a, per unit of time.

b^l : fleet size (total number of buses) of line l.

$\overline{B}$: total number of buses in the system.

$\overline{c}_a$: travel time spent by a transit vehicle through link a.

g_a^l : operating cost experienced by a bus of line l going through link a.

$\{\overline{\delta}_{al}\}$: link-line incidence matrix:

-takes value 1, if line l passes through link a,

-and 0 otherwise.

We also define the following vectors:

$\overline{f} = (\overline{f}^l)_{l\in\mathcal{L}}$: vector of transit flows.

$B = (b^l)_{l\in\mathcal{L}}$: vector of lines fleets.

$F = (f, \overline{f})$: multimodal flow vector .

The following relationships are direct consequences of the above definitions:

$$\begin{aligned} \overline{f}_a &= \sum_{l\in\mathcal{L}} \overline{\delta}_{al}\overline{f}_{al} \quad \forall a \in \mathcal{A} \\ \overline{f}^l &= \sum_{a\in\mathcal{A}} \overline{\delta}_{al}\widetilde{f}^l_a \quad \forall l \in \mathcal{L} \\ \varphi_a &= \sum_{l\in\mathcal{L}} \overline{\delta}_{al}\varphi^l \quad \forall a \in \mathcal{A} \\ \overline{B} &= \sum_{l\in\mathcal{L}} b^l \quad . \end{aligned} \tag{3}$$

As was explained previously, cars and buses simultaneously share the same infrastructure (city streets), and therefore their respective congestion functions are dependent on both car and bus flows. We assume that bus travel times and operating costs on link a will also be influenced by the total number of passengers boarding on link a :

$$c_a(f_a, \varphi_a) = \alpha_a + \beta_a(f_a + \mu\varphi_a)^m, \tag{4}$$

$$\overline{c}_a(f_a, \varphi_a) = \overline{\alpha} + \overline{\beta}(f_a + \mu\varphi_a)^{\overline{m}} + \gamma_a \sum_{l\in\mathcal{L}} \widetilde{f}^l_a, \tag{5}$$

$$g_a(f_a, \varphi_a) = \widetilde{\alpha} + \widetilde{\beta}(f_a + \mu\varphi_a)^{\tilde{m}} + \widetilde{\gamma}_a \sum_{l\in\mathcal{L}} \widetilde{f}^l_a, \tag{6}$$

We also assume that car drivers and transit system users strive to minimize travel time (or "generalized" travel costs) while bus operators try to maximize individual profits. Therefore, car drivers behave according to Wardrop's first principle of traffic equilibrium (user-optimized flow pattern). To determine how the passengers utilize the services provided over the transit network we will use state-of-the-art transit assignment models based on the concept of a "strategy" (see Spiess [83]) or, equivalently, "hyperpath" (Nguyen and Pallottino [85]), or

"transit route" (De Cea and Chapleau [84]). In these models, a passenger waiting at a bus stop selects the first incoming bus from a set of "attractive lines". This set can be chosen such that the passenger's expected travel time (including waiting time) is minimized. The models differ in the number of lines that can be members of the set of attractive lines. For each individual passenger the actual path followed is random (bus arrivals are random) while the strategy is deterministic.

We assume that a system equilibrium is achieved when no player (bus operator, car user or transit passenger) can improve his objective by an unilateral modification of his strategy. In particular the following conditions will be fulfilled at equilibrium:

(i) No car driver can reduce his travel time by a unilateral change in the path used.

(ii) No change of strategy can decrease expected travel time experienced by transit users.

(iii) No bus operator can increase his profit by a unilateral change in the line served by his vehicle.

This global scheme represents a generalization of the classical Cournot-Nash equilibrium concept where, instead of two profit-maximizing players, we consider two systems that must achieve an equilibrium among and within themselves. In the particular instance where the equilibrium state of each system corresponds to the global maximum of some objective function, then an overall equilibrium corresponds to a Cournot-Nash equilibrium of the corresponding two-player game. An equilibrium will be identified by a set (F^*, φ^*) of multimodal equilibrium flows, and profit maximizing frequencies.

It is possible to make different assumptions with respect to the degree of deregulation in the transit system. Full, free competitive conditions would be

obtained if entry into the market was unrestricted and fares were determined from the sole interaction of supply and demand. Nevertheless some intermediate situations can also be considered and modelized. In this paper we will assume that the fare structure is exogenously determined. Even in this situation some important questions have to be resolved: Firstly, how many buses will be operating in the system? And secondly, what will be their distribution on the network of transit lines? The model presented in this paper will address those two questions as a first approximation to the solution of a more general model that considers full free competitive conditions. Some alternative formulations, corresponding to different market conditions, will be mentioned in the last section of the paper. In the sequel we assume that passenger flows and/or strategies are provided by some technologically up-to-date transit assignment model such as EMME/2 (see Babin et al. [82]) or MADITUC (see De Cea and Chapleau [84]).

3. Operators equilibrium model formulation.

3.1 Equilibrium conditions for the transit market.

Given that transit services are provided by a large number of individual operators who strive to maximize individual profits, and that they freely choose the lines where to run their buses, the equilibrium model can be formulated as an assignment-of-buses-to-lines problem. The question of interest is how to assign a given fleet of buses to the set of potential lines $\mathcal{L}$ such that individual operators profits are indeed maximized.

According to our assumptions, an equilibrium will be achieved when no operator can improve his profit by a unilateral change of line. Therefore at equilibrium, all lines served must produce the same profit per vehicle and time unit, and this profit must not be lower than the profit that would be obtained by operating a vehicle on any of the unserved lines.

These conditions can be expressed mathematically as follows:

$$\Pi_l \begin{cases} = u, & \text{for all } l \in \mathcal{L}, \text{ such that: } \quad \varphi^l > 0, \quad b_l > 0, \\ \leq u, & \text{for all } l \in \mathcal{L}, \text{ such that: } \quad \varphi^l = 0, \quad b_l = 0, \end{cases} \tag{7}$$

where Π_l denotes the net profit obtained by a vehicle operating on line l and u represents the maximum net revenue obtained, per vehicle and per time unit, on all served transit lines, at equilibrium.

Under the assumptions made so far we propose two formulations for the problem. In the first formulation the total number of buses in the system is fixed and we want to determine their distribution on the set of potential lines of the transit network such that individual profits are maximized (operators equilibrium). Then, given fixed fleet size $\overline{B}$ and O-D transit demands, under or overcapacity in the system cannot be ruled out a priori. Furthermore, since the fare structure is exogenously given, equilibrium profits will in general be either positive ($u > 0$) or negative ($u < 0$), depending on the magnitudes of demand, fleet size, and fare values.

The second formulation considers not only the freedom for existing operators to choose the lines they operate, but also free entry and exit from the transit market. The perspective of a positive profit would attract new operators into the the system, therefore driving benefits to zero while negative profits would induce some operators to leave, increasing the benefits of those staying in the system. Consequently, in this case, the equilibrium conditions (7) must hold with u equal to zero.

It is interesting to note that the profit u can be interpreted as the Lagrange multiplier (dual variable) associated with the total fleet constraint.

It is easy to show (Fernandez and Bravo [86]) that conditions (7) can be expressed as the following variational inequality:

$$\Pi(B^*, F^*)(B^* - B) \geq 0, \qquad \forall B \in \Omega, \tag{8}$$

where Π represents the row vector $(\Pi_l)_{l\in\mathcal{L}}$, B is the column vector $(b_l)_{l\in\mathcal{L}}$ and Ω the feasible region defined by the constraints:

$$\sum_{l\in\mathcal{L}} b_l = B, \qquad b_l \geq 0.$$

Finally we observe that the variational inequality (8) is equivalent to the nonlinear complementarity problem:

$$\begin{aligned} (\Pi(B,F) - u) &\leq 0 \\ (\Pi(B,F) - u)\cdot B &= 0 \\ B &\geq 0 \\ \sum_{l\in\mathcal{L}} b_l &= \overline{B}. \end{aligned} \tag{9}$$

If the profit function has a symmetric Jacobian throughout the feasible region Ω then formulations (8) and (9) are satisfied by any critical point of the optimization problem:

$$\begin{aligned} \max_{B\in\Omega} \quad & \int_0^B \Pi(x,F)\,dx \\ \text{s.t.} \quad & \sum_{l\in\mathcal{L}} b_l = \overline{B}, \qquad B \geq 0. \end{aligned} \tag{10}$$

For the second formulation, in which free entry and exit form the transit market is assumed, the total fleet constraint is dropped.

3.2 Bus operators profit function.

The net profit obtained by a bus operator will be the consequence of the total income received from fares minus the operating costs incurred. Both revenues and costs are related to the number of round trips made per time unit. Therefore the first step in defining a profit function consists in devising an expression for bus frequencies, considering congestion effects. Let t_l denote the round trip travel time on line l. We can write:

$$t_l = \sum_{a\in\mathcal{A}} \overline{\delta}_{al}\overline{c}_a(f_a,\varphi_a).$$

It is important to notice that t_l is a function of the entire frequency vector, i.e.:

$$t_l = t_l(\Phi)$$

where $\Phi = (\varphi_l)_{l \in \mathcal{L}}$, and the frequency on line l is given by:

$$\varphi_l = h_l(\Phi) \stackrel{\text{def}}{=} b_l / t_l(\Phi). \tag{11}$$

Consequently the frequency vector Φ will be solution to the fixed point problem:

$$\Phi = H(\Phi) \tag{12}$$

with H defined as the vector $(h_l)_{l \in \mathcal{L}}$.

Recalling that $\overline{f}^l$ denotes the total number of passengers boarding line l per unit of time, and assuming that the fare (by trip) is constant on all lines and has value ρ, then the net profit obtained by each vehicle operating on line l, per time unit, is given by:

$$\Pi_l(B, F) = (\rho \overline{f}^l / b_l) - (\sum_{a \in \mathcal{A}} \overline{\delta}_{al} g_a^l)/t_l - FC_l, \tag{13}$$

where FC_l denotes the fixed cost of operating a vehicle on line l, per unit of time. This expression can be somewhat simplified if we assume that the variable operating cost g_a^l is proportional to the time spent on the network, i.e. that increased fuel consumption due to heavier traffic depends linearly on travel time, and not on the distance travelled. In this case the variable operating cost per round trip g_l is proportional to the round trip travel time t_l, i.e.: $g_l = k\, t_l$ for some positive constant k and the expression of the profit function reduces to:

$$\Pi_l(B, F) = (\rho \overline{f}^l / b_l) - K_l, \tag{14}$$

where $K_l \stackrel{\text{def}}{=} k - FC_l$.

4. Solution approaches.

4.1 Finding the users-operators equilibrium.

We will now present two solution approaches for the transit market equilibrium problem (TMEP) formulated in section 3. As was mentioned in section 2, the TMEP structure corresponds to an extension of a non cooperative game between two systems of players. Transit operators strive to maximize their operating profits, which are affected by the behaviour of system users, and commuters are bound by the decisions of the transit operators, minimizing their objective (transit time) accordingly.

Given that the equilibrium conditions of the system users can also be expressed using the variational inequality formalism, the combined users-operators problem can be expressed as the system of variational inequalities:

$$\begin{aligned} \Pi(B^*, F^*)(B^* - B) \geq 0 \qquad & \forall B \in \Omega \\ C(B^*, F^*)(F^* - F) \leq 0 \qquad & \forall F \in \Gamma \end{aligned} \tag{15}$$

where Γ denotes the set of feasible flow patterns, i.e. the set of nonnegative vectors satisfying flow conservation and demand constraints. This system can be solved by means of iterative methods such as the block-Jacobi or block-Gauss-Seidel schemes, where, at each iteration, an equilibrium problem involving one group of players (users or operators) is solved, given that the "strategy" of the other group is temporarily fixed to its previous value. The iterative process is stopped when both equilibria are consistent. The Gauss-Seidel scheme possesses the advantage of using the latest information available, and will generally be preferred.

Given that efficient algorithms exist for the resolution of the passengers assignment problem (see Florian and Spiess [83]) we will concentrate on solution approaches for the operators equilibrium problem.

Since a block Gauss-Seidel iterative algorithm will be used to solve the system of variational inequalities (15), some information has to be transferred between the subsystems, at a each iteration. This information will describe the transitory equilibrium reached by the last subsystem, given the current state of the other subsystem. A passenger equilibrium is described in terms of equilibrium flows, observed over the current transit network. This information is then used to calculate operators profits in equation (13).

In the following section we present two different approaches to obtain an equilibrium for the operators subsystem. They differ in the information structure transferred from the passengers subsystem and in the way this information is utilized to compute operators profits.

4.2 First approach.

In this solution approach the information that is transferred from the passengers to the operators subsystem, at each iteration, corresponds to passenger loads observed on each line at that precise instant. Vehicles will be assigned to the different transit lines, trying to reach equal profits on all lines served, based on profits calculated using the current information about passenger loads. Therefore the consequences of the changes in boardings resulting from frequency changes are not anticipated by the bus operators. They will only be taken into account at the next iteration, after a new passenger equilibrium has been computed.

Consequently, at each iteration, the operators equilibrium problem is solved using profit function formulation (13) or (14), with $\overline{f}^l$ fixed and equal to the total number of transit users assigned to line l, at the previous iteration of the passenger equilibrium. Then $\overline{f}^l$ will be equal to:

$$\overline{f}^l = \sum_{a \in A} \sum_{j \in \mathcal{N}} \widetilde{f}^l_{aj} \quad , \tag{16}$$

where the $\widetilde{f}^l_{aj}$ are constant and obtained from the last transit assignment per-

formed.

This algorithmic approach has the limitation that if no passenger is assigned on a line at some iteration of the iterative procedure, then no vehicles will be assigned at any of the subsequent iterations, given that operators react only to observed (not predicted) flows. Therefore, in order that the procedure works properly, it is necessary to start from a point in Ω for which $b_l > 0, \forall l \in \mathcal{L}$. Otherwise some lines might be abandoned at the beginning of the algorithmic process, never to reappear later on. A line will disappear from consideration as soon as a zero frequency is assigned to it. This can be a consequency of a zero load assigned by the users model at the last iteration, or the line being non-profitable, even though demand for that service exists. This will occur when revenues are not sufficient to cover the sum of fix and operating costs.

4.3 Second approach.

This approach tries to overcome the limitations of the previous decomposition scheme. It uses a more sophisticated algorithm that is based on the fact that in a transit assignment model using the strategy concept, strategies represent the primary variables, from which flows can be easily deduced. We will therefore assume that bus operators react to strategies rather than to actual flows, i.e. that they can anticipate the changes in passenger loads resulting from changes in frequencies.

Let us denote by U_{aj} the set of transit lines attractive to passengers boarding at arc a and having final destination j, i.e. that those passengers will board the first incoming bus belonging to the set of lines U_{aj} (this set is a byproduct of the transit assignment model). Given the probabilistic assumptions underlying the transit assignment model (exponential headways for buses), the expected number of boardings will be proportional to frequencies on attractive lines. This statement will hold true until the set U_{aj} changes, due to significant changes

in bus frequencies or headways; for instance, if frequency on the quickest route becomes very high, then the set of attractive lines will eventually shrink to that single line. Broadly speaking, when the actual frequencies are input back into the transit assignment model, this might lead to changes in the sets U_{aj}.

Therefore, in this formulation, the profit function for the operators subproblem (13) or (14) will use the value of $\overline{f}^l$ given by the following expression:

$$\overline{f}^l = \sum_{a \in \mathcal{A}} \sum_{j \in \mathcal{N}} \frac{\varphi_l}{\sum_{k \in U_{aj}} \varphi_k} \overline{\delta}_{al} \overline{\delta}_{jl} \overline{f}_{aj} \quad , \qquad \forall l \in U_{aj}, \tag{17}$$

where $\overline{f}_{aj}$ is the total number of travellers going through nodes a and j at the current iteration. It is important to note that if $l \notin U_{aj}$, then $\overline{f}^l = 0$. In this approach, at each iteration of the operators subproblem, the data U_{aj} and $\overline{f}_{aj}$ are held fixed. An important feature of this formulation is that, although a line may be assigned zero frequency at a given iteration, its frequency can be raised at a later iteration. Assume for instance that the quickest line between an origin and a destination has been assigned zero frequency at the start of the calculations. Although the flow on this line will be zero at the first iteration, it will nevertheless be part of the set of attractive lines connecting the given O-D pair, and will obviously be reactivated at the next iteration, if the expected (non zero) demand is sufficient to yield nonnegative profits on this line.

It is straightforward to see that both approaches will yield equivalent solutions if the respective sets of lines with nonzero frequencies are the same. Otherwise it might be that the solutions differ. A simple example illustrating both approaches is presented in the Appendix.

5. Implementation.

The first approach is obviously the easiest to implement, not requiring the information about passengers strategies. However it suffers two drawbacks:

(i) it can be stuck in a solution that does not have the right set of active lines (lines with nonzero frequencies)

(ii) since less information is taken into account in the operators subproblem (first approach) it is expected that global convergence of the algorithm could be slow and, in some extreme cases, the algorithm could cycle. Numerical results illustrating the application of both approaches to a realistic transit network (Santiago, Chile) will be presented in a forthcoming paper.

6. Existence and uniqueness.

The existence of at least one equilibrium solution to the combined users-operators model follows, by a fixed point theorem, from the continuity of the functions involved and the compactness of the set of feasible solutions. The latter condition is automatically satisfied since demand is assumed exogenous and arbitrarily large frequencies would yield, in the free entry case, negative profits, clearly a non-equilibrium situation.

The uniqueness issue is more problematic. The model cannot distinguish between perfectly matched lines (identical sequence of arcs, stops, transfer nodes). Whereas uniqueness results based on strict monotonicity arguments can be obtained for the passengers submodel as well as the operators submodel in the first formulation, it is less obvious for the operators submodel in the second formulation, let alone the combined model. However partial results can be obtained under fairly weak assumptions in the free entry case. It must be reminded however that strict monotonicity is a sufficient but not necessary condition for uniqueness of the equilibrium solution.

For the combined users-operators model however, uniqueness in the first formulation is certainly not satisfied since any line assigned a zero frequency in the first iteration will be excluded from the market, however attractive it might be. Therefore uniqueness is probably best addressed by investigating the second for-

mulation. Whereas it cannot be expected to theoretically prove uniqueness in all cases, we feel that uniqueness should be achieved in most interesting practical situations. This statement can of course only be substantiated by thorough numerical experimentation (see the appendix for an example where the solution is not unique).

7. Alternative formulations of the operators market.

Alternative formulation of the operators equilibrium problem can be obtained if different market conditions are assumed. If we assume that only one private operator controls the whole transit system, it will behave as a monopoly trying to maximize total profit generated by the transit system operation. In such case an equilibrium solution will be obtained by solving the following bilevel programming problem:

$$\begin{aligned} \max_{\{B\}} \quad & \Pi(B,F) \\ \text{s.t.} \quad & F: \quad \text{bimodal} - \text{equilibrium}, \\ & b_l \geq 0, \qquad \forall l \in \mathcal{L} \quad . \end{aligned} \tag{18}$$

Here Π represents the sum of the profits on all lines and vehicles of the system. Profits per vehicle will still be given by expression (13).

Problem (18) can be solved using an approach similar to the one used in section 4, based on the following economic observation: "In order that the total profit over the entire system be maximized, all lines should yield equal marginal benefits". Therefore conditions (7), (8) and (9) can be reformulated in terms of marginal instead of average profits, and similar solution algorithms can be implemented.

When Π in (18) represents social benefit instead of private profits, then the problem corresponds to a bilevel programming problem, or in this particular case, a transit design problem (see Fernandez [86]). Although this is considered to be

an intractable problem in its general form, efficient heuristic procedures have been design to solve particular instances, such as the equilibrium network design problem (see Marcotte, [86]).

An alternative formulation is obtained if we assume that each line $l \in \mathcal{L}$ is operated by a single transit firm. Then, in the operators subsystem, there will be as many players as the number of lines operated. Each player will strive to maximize its own total benefit, without explicit knowledge of the other players profit functions. An equilibrium will be reached when each player is at a maximum, given the current decisions of all other players. Then an equilibrium solution can be computed by simultaneously solving all operators optimization problems (Nash equilibrium, see Friedman [77]):

$$\begin{aligned} \max_{\{B\}} \quad & \Pi_l(F, B) \qquad \forall l \in \mathcal{L} \\ \text{s.t.} \quad & F: \quad \text{bimodal} - \text{equilibrium} \\ & b_l \geq 0, \qquad \forall l \in \mathcal{L} \quad . \end{aligned} \tag{19}$$

For the solution of this problem the Jacobi iterative scheme seems appropriate. At each iteration, players maximize their own profits given the other players decision at the previous iteration.

REFERENCES

A. Babin, M. Florian, L. James-Lefevre and H. Spiess, "EMME/2 : Interactive Graphic Method for Road and Transit Planning", *Transportation Research Record* **866** 1–9 (1982).

F.D. Bard and J.E. Falk, "An explicit solution to the multilevel programming problem", *Computers and Operations Research* **9** 77–100 (1982).

J. De Cea and R. Chapleau, "MADITUC : Un modelo de asignación de rutas multiples en redes de transporte público", *Apuntes de Ingeniera* **15** 113–140

(1984).

J.E. Fernandez and F. Bravo, "Modelación de sistemas de transporte público urbano", III CLAIO, 18–22 Agosto, Santiago, Chili (1986).

J.E. Fernandez, "Transit networks: Simulation vs. Optimization approaches", 14th PTRC Summer Annual Meeting, University of Sussex, Brighton, England (1986).

M. Florian and H. Spiess, "On binary mode choice/assignment models", *Transportation Science* **17** 32–47 (1983).

J.W. Friedman, *Oligopoly and the theory of games*, North-Holland, Amsterdam (1977).

P. Marcotte,"Continuous Network Design Problem with Congestion Effects: a case of Bilevel Programming", *Mathematical Programming* **34**, 142-162 (1986).

S. Nguyen and S. Pallotino, "Equilibrium traffic assignment for large scale transit networks", Quaderno **14**, Istituto per le applicazione del calcolo Mauro Picone, Consiglio Nazionale delle Ricerche, Rome, Italy (1985).

H. Spiess, "On optimal route choice strategies in transit networks", publication **286**, Centre de Recherche sur les Transports", Université de Montréal, Montréal, Canada (1983).

APPENDIX

NUMERICAL EXAMPLE

Consider a transit system satisfying the following assumptions:

- Exponential headways for buses. Therefore, expected waiting times will be equal to headways.
- Passengers use transit services such that expected travel times are minimized. They select the optimum "strategy" to travel between each pair of nodes, giving a zero weight to travel costs. (1)
- Fixed flat fares of 10 monetary units on all lines
- Flows will be considered only in one direction (no return flows). Therefore we will assume that once a bus reaches the last node of the line it instantaneously reappears at the first node; Thus, round trip times will be equal to one-way travel times between the first and last nodes of the line.

Figure 1 illustrates the transit network used for our example; travel times are indicated in parenthesis for each line. Operating costs per unit of time will be assumed fixed and equal to 120, 150 and 200 monetary units, for each vehicle, on lines 1, 2, and 3 respectively. Travel demand is set to 120, 60 and 80 (trips per unit of time) between O-D pairs A-C, A-B and B-C respectively.

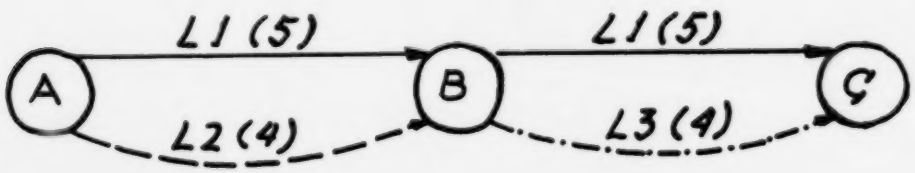

Fig. 1.

(1) This assumption would be unrealistic in the context of transportation in developing countries, where time has in general a low monetary value. Nevertheless it does not affect the results we want to illustrate here and is only made in order to simplify calculations.

The initial state of the system is given by:

$$B = (10, 2, 1), \qquad \varphi = (1, 1/2, 1/4).$$

In this simple case it is not difficult to see that all three lines are included in the optimum strategy (all of them are attractive). To travel between any of the three O-D nodes with positive demands, passengers will board the first incoming vehicle (from line 1 or 2 if they leave node A, or from line 1 or 3 if they leave node B).

Therefore boardings will be split at node A with 2/3 of the trips assigned to line 1 and 1/3 to line 2. At node B, 4/5 of the trips will board line 1 and 1/5 line 3.

Loads on each line are given in Table 1 below.

LINES	SECTION A-B		SECTION B-C	
LOADS	Line 1	Line 2	Line 1	Line 3
Pass. from A to C	80	40	80	
Pass. from A to B	40	20		
Pass. from B to C			64	16
Pass. transf. at B			32	8
TOTAL	120	60	176	24

Table 1. Initial solution: loadings.

Expected transit times ($\bar{t}_{i,j}$) are given by: [2]

$$\bar{t}_{A,B} = \frac{2}{3}\left(\frac{2}{3}+5\right) + \frac{1}{3}\left(\frac{2}{3}+4\right)$$

$$\bar{t}_{B,C} = \frac{4}{5}\left(\frac{4}{5}+5\right) + \frac{1}{5}\left(\frac{4}{5}+4\right)$$

$$\bar{t}_{A,C} = \frac{2}{3}\left(\frac{2}{3}+5\right) + \frac{1}{3}\left[\frac{2}{3}+4+\frac{4}{5}\left(\frac{4}{5}+5\right)+\frac{1}{5}\left(\frac{4}{5}+4\right)\right]$$

[2] The interested reader is referred to Spiess [83] or Nguyen and Pallottino [85], for details of the calculations.

or:

$$\bar{t}_{A,B} = 5.33, \quad \bar{t}_{B,C} = 5.60 \quad \text{and} \quad \bar{t}_{A,C} = 10.53$$

Equilibrium solutions

In this case it is not difficult to show that there are two distinct equilibrium solutions for the users/operators system. The first one corresponds to:

$$B = (0, 12, 10), \qquad \varphi = (0, 3, 2.5)$$

and the second one to:

$$B = (20, 4, 0), \qquad \varphi = (2, 1, 0)$$

The latter solution corresponds to a degenerate situation where strategies are not unique. Indeed, passengers going from A to B do not modify their expected transit times by including (or not) line 1 in their strategies. Therefore, there will be many different possible equilibrium loads between A and B when the line frequencies are those given by the second solution; however, only one will be compatible with a combined users-operators equilibrium. It corresponds to that equilibrium where boardings are split between lines 1 and 2 in proportion to their frequencies. Then, 2/3 of trips will take line 1 and 1/3, line 2. However, at node B everybody will have to board line 1, given that it is the only one with a positive frequency at that node. It is interesting to note that if passengers choose any of the other optimum strategies possible at node A, a combined users- operators equilibrium does not exist. However, a "small" data perturbation will make the strategies unique, (i.e. reduce by an arbitrary small amount travel time on line 1).

Line loads and travel times for the equilibrium solutions are given below.

a) First Solution: $B = (0, 12, 10)$

Transit times will be: $\bar{t}_{A,B} = (1/3) + 4 = 4.33$,

$\bar{t}_{B,C} = (2/5) + 4 = 4.40$ and $\bar{t}_{A,C} = 4.33 + 4.40 = 8.73$.

Loadings are given in Table 2.

LINES	SECTION A-B		SECTION B-C	
LOADS	Line 1	Line 2	Line 1	Line 3
Pass. from A to C		120		
Pass. from A to B		60		
Pass. from B to C				120
Pass. transf. at B				80
TOTAL		180		200

Table 2. First Equilibrium Solution: loadings.

b) Second Solution: $B = (20, 4, 0)$

Transit times are equal to:

$$\bar{t}_{A,B} = \frac{2}{3}\left(\frac{1}{3} + 5\right) + \frac{1}{3}\left(\frac{1}{3} + 4\right) = 5.00$$

$$\bar{t}_{A,C} = \frac{1}{2} + 5 = 5.50$$

$$\bar{t}_{A,C} = \frac{2}{3}\left(\frac{1}{3} + 10\right) + \frac{1}{3}\left(\frac{1}{3} + 4 + \frac{1}{2} + 5\right) = 10.17$$

Loadings are given in Table 3:

LINES	SECTION A-B		SECTION B-C	
LOADS	Line 1	Line 2	Line 1	Line 3
Pass. from A to C	80	40	80	
Pass. from A to B	40	20		
Pass. from B to C			80	
Pass. transf. at B			40	
TOTAL	120	60	200	

Table 3. Second Equilibrium Solution: loadings.

We can check that the above solutions are in equilibrium, by computing the bus operators profits in each case. For the first solution we obtain:

$$\Pi_2 = \left(\frac{180}{12}\right) 10 - 150 = \quad 0, \qquad \Pi_3 = \left(\frac{200}{10}\right) 10 - 200 = \quad 0,$$

and we can use (17) in the paper to evaluate the potential profits on line 1, which in this case will be $\Pi_1 = -120$. Considering that line 1 is not on the attractive sets $U_{A,B}, U_{A,C}$, there will be no income from fares.

Therefore, profits are zero on lines served and potential profits are negative on the line not served (line 1), which means that the system is at equilibrium.

Similarly, for the second solution:

$$\Pi_1 = \left(\frac{240}{20}\right) 10 - 120 = \quad 0, \qquad \Pi_2 = \left(\frac{60}{4}\right) 10 - 150 = \quad 0,$$

and using formula (17) to evaluate the potential profits on line 3:

$$\Pi_3 = 120 \cdot \left(\frac{1/4}{2 + 0/4}\right) - 200 = \quad -50,$$

which is non-positive, as it should be.

Equilibrium calculation

Starting from our initial solution, $B = (10, 2, 1)$, we will illustrate one complete iteration of the algorithms proposed in this paper.

1) First algorithmic approach:

In accordance with Table 1 data, the current profits on each line will be given by:

$$\Pi_1 = \left(\frac{216}{10}\right) 10 - 120 = \quad 96, \qquad \Pi_2 = \left(\frac{60}{2}\right) 10 - 150 = \quad 150,$$

$$\Pi_3 = \left(\frac{24}{1}\right) 10 - 200 = \quad 40.$$

According to the first approach, we assume that flow division among competing lines stays constant, and the number of buses corresponding to zero profits on all three lines is obtained by solving the following set of equations:

$$\left(\frac{2160}{B_1}\right) - 120 = \left(\frac{600}{B_2}\right) - 150 = \left(\frac{240}{B_3}\right) - 200 = \quad 0,$$

whose solution is: $B = (18, 4, 1.2)$

Then, at the users level we recalculate the strategies, and obtain a new transit assignment solution, which in turn will be used to compute a new operators equilibrium. This process is repeated until a simultaneous users/operators equilibrium is achieved. In this case five iterations are necessary to find such an equilibrium. The solutions obtained at each iteration are the following:

Iteration 1: $B^{(1)} = (18.0, 4.0, 1.2)$,

Iteration 2: $B^{(2)} = (14.3, 12.0, 1.5)$,

Iteration 3: $B^{(3)} = (13.3, 12.0, 2.0)$,

Iteration 4: $B^{(4)} = (12.2, 12.0, 2.7)$,

Iteration 5: $B^{(5)} = (0, 12, 10)$.

At the second iteration, line 1 is dropped from the optimal strategy on section A-B resulting in line 2 increasing considerably its market share. Later on, at iteration 5, line 1 disappears from the optimal strategy of section B-C, leading to the first equilibrium solution. To see how the number of iterations required to find an equilibrium depends on the data, observe that only two iterations would be required if the cost on line 3 was reduced or cost on line 1 is increased.

2) Second algorithm: In this approach, profits per bus are given by:

$$\Pi_1 = \left(\frac{180 \times 10}{B_1}\right)\left(\frac{B_1/10}{B_1/10 + B_2/4}\right) + \left(\frac{120 \times 10}{B_2}\right)\left(\frac{B_1/10}{B_1/10 + B_3/4}\right) - 120,$$

$$\Pi_2 = \left(\frac{180 \times 10}{B_2}\right)\left(\frac{B_2/4}{B_1/10 + B_2/4}\right) - 150,$$

$$\Pi_3 = \left(\frac{120 \times 10}{B_3}\right)\left(\frac{B_3/4}{B_1/10 + B_3/4}\right) - 200.$$

and after setting: $\varphi_1 = B_1/10$, $\varphi_2 = B_2/4$ and $\varphi_3 = B_3/4$, the following nonlinear complementarity problem is obtained:

(NLCP):

$$(\Pi_1) \qquad \left(\frac{180}{\varphi_1 + \varphi_2}\right) + \left(\frac{120}{\varphi_1 + \varphi_3}\right) - 120 \leq 0,$$

$$(\Pi_2) \qquad \left(\frac{450}{\varphi_1 + \varphi_2}\right) - 150 \leq 0,$$

$$(\Pi_3) \qquad \left(\frac{300}{\varphi_1 + \varphi_3}\right) - 200 \leq 0,$$

$$(\varphi) \qquad \varphi_1, \quad \varphi_2, \quad \varphi_3 \quad \geq 0,$$

$$(\Pi \cdot \varphi) \qquad \Pi_1 \cdot \varphi_1 \quad = \quad \Pi_2 \cdot \varphi_2 \quad = \quad \Pi_3 \cdot \varphi_3 \quad = 0,$$

whose unique solution is: $\varphi_1 = 2$, $\varphi_2 = 1$, $\varphi_3 = 0$, corresponding to the second equilibrium solution: $B_1 = 20$, $B_2 = 4$, $B_3 = 0$.

Remark. Consider the operating cost on line 3 as a parameter and let its value be denoted by c. For c larger than 150, the solution is not modified. For $c = 150$, problem (NLCP) admits a continuum of solutions. Finally, for c less than 150, line 1 is driven out of the market, and only one equilibrium solution is obtained (both methods yield the same solution). For instance, for $c = 100$, the unique equilibrium solution will be: $B_1 = 0$, $B_2 = B_3 = 12$.